|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Ans** : Total outcomes = 8

i.e., {HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}

So Probability of 2 heads and 1 tail / Total outcomes = 3/8.

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

**Ans**: If two Dice are rolled ,then the possible outcomes are 36.

Favourable outcomes(i.e. equal to 1) = **0**

because, they starts with the probability (1,1)  and the sum may not be  equal

to 1.

1. Less than or equal to 4

**Ans** :There are 6 \* 6 possible outcomes,i.e.  = 36

Of these there are 6 totals which are less than or equal to 4.

i.e. P(less than or equal to 4) = 6/36 = 1/6.

c)Sum is divisible by 2and 3

**Ans** :When 2 dice are rolled the possible outcomes are :

(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)

Therefore, total number of possible outcomes = 36.

So there are 24 sum of totals which are divisible by 2 and 3.

Thus the probability = 24/36 = 2/3.

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

**Ans:** There are7 balls in total. From these 5 balls are not in blue.

So, we want to select favorable cases as 2 balls from those 5 balls.

Similarly for total cases i.e., 2 ball from 7 balls.

Probability = 5C2/7C2 = [(5\*4/1\*2)/ (7\*6/1\*2) = 10/21

So, the probability of none of the balls are blue = 10/21 ways.

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Ans:** To calculate expected no of candies we have, **∑**X. P(X).

So, (1\*0.015) +(4\*0.20) +(3\*0.65) +(5\*0.005) +(6\*0.01) +(2\*0.120)

Expected no of candies for randomly selected child = 3.09.

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
| **Ans:** Please check Ans\_7. |  |  |  |

Q8) Calculate Expected Value for the problem below

(a)The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**Ans:** To find Expected Value we have,

Expected Value = ∑X. P(X)

There are 9 patients in the clinic. So, the probability of choosing one at random is 1/9. So, for all the patients expected and probability is:

108, 110,123,134,135,145,167,187,199 = 1/9

Expected Value = (1/9)\*180 + (1/9)\*110 + (1/9)\*123 + (1/9)\*134 + (1/9)\*135 + (1/9)\*145 + (1/9)\*167 + (1/9)\*187 + (1/9)\*199

= 1/9 (180+110+123+134+135+145+167+187+199)

=1/9 \* 1308

=145.3333

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :** Please Check Ans\_9(a).

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** Please Check Ans\_9(b).

**Q10) Draw inferences about the following boxplot & histogram**



**Ans:** Histogram:

1.ChickWeight$weight is Right skewed or positive skewed.

2.Most of the ChickWeight$weight Frequency is between 50 – 150.

3.The Most of the points are concentrated in the range 50-100 with frequency 200.



**Ans: BoxPlot:**

1.The data is Right Skewed.

2.There are outliers in the upper side.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Ans:** Please check Ans\_11.

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median,variance,standard deviation.
2. What can we say about the student marks?

**Ans:** Please check Ans\_12.

Q13) What is the nature of skewness when mean, median of data is equal?

**Ans :** If the distribution is symmetric, then mean and median are equal ,and distribution has zero skewness.ie Data is normalized and there is no skewness.

Q14) What is the nature of skewness when mean >median?

**Ans :** If mean is greater than median, the distribution is positively skewed(or right skewed.

Q15) What is the nature of skewness when median > mean?

**Ans:** When median is greater than mean then the distribution is negatively skewed (or left skewed).

Q16) What does positive kurtosis value indicates for a data?

**Ans:** Positive values of kurtosis indicate that the distribution is peaked and possess thick tails.

Q17) What does negative kurtosis value indicates for a data?

**Ans:** Negative values of kurtosis indicates that the distribution is flat and has thin tails.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**Ans:** The data distribution is not normally distributed or it is asymmetric in nature.

And it is

What is nature of skewness of the data?

**Ans:** The data possess a left skew distribution.

What will be the IQR of the data (approximately)?

**Ans:** To find IQR from the box plot distribution, we have

IQR = Q3 – Q10

= 18 – 10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Ans:** 1. Both the Boxplots, i.e. Boxplot 1 and Boxplot 2 are normally distributed.

2. Median of boxplot 1 and Boxplot 2 are same.

3. The dispersion of data is more in Boxplot 2 than Boxplot1.

4. In Boxplot1, data points consistently hover around the center values. But in Boxplot2, there are more variable.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG<- Cars $ MPG

* 1. P(MPG>38)

**Ans:**

Mean of MPG = 34.422

Std Dev of MPG = 9.13144

For getting probabilities of normal distribution with μ and σ :

Pnorm(x , mean = ,std = ,lower.Tail = )

For P(MPG>38) = 1 – pnorm(38, 34.422, 9.13144)

= 0.3475

* 1. P(MPG<40)

**Ans:** P (MPG<40) = pnorm(40, 34.422, 9.13144)

=0.7293

* 1. P (20<MPG<50)

**Ans:** P (20<MPG<50) = pnorm(50, 34.422, 9.13144) – (1 – pnorm(20, 34.422, 9.13144))

= 0.0131

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** Please check Ans\_21(a)

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) fromwc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Ans:** Please check Ans\_21(b)

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans:** Please check Ans\_22.

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Ans :** Please check Ans\_23.

Q 24**)**A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom

**Ans:** Please check Ans\_24.